



Mobile phone application data for activity plan generation

Downloaded from: <https://research.chalmers.se>, 2026-10-30 19:32 UTC

Citation for the original published paper (version of record):

Tozluoglu, Ç., Liao, Y., Sprei, F. (2026). Mobile phone application data for activity plan generation. *Transportation*, 53(7): 5121-5153. <http://dx.doi.org/10.1007/s11116-025-10679-6>

N.B. When citing this work, cite the original published paper.



Mobile phone application data for activity plan generation

Çağlar Tozluoğlu¹ · Yuan Liao^{1,2} · Frances Sprei¹

Accepted: 27 August 2025 / Published online: 20 September 2025
© The Author(s) 2025

Abstract

Activity-based models in transport are crucial for providing a comprehensive and realistic understanding of individuals' activity-travel patterns, improving demand forecasting, policy analysis, and land use and transportation planning. While travel surveys have long served to develop activity-based models with complete activity-travel plans, they are often costly to collect and have small sample sizes. Mobile phone application data, one example of emerging mobility data sources, offers an alternative with broader geographical and population coverage over extended periods, which remains under-exploited in activity-based models. However, the challenges of using these data include sampling biases in the population coverage and data sparsity at the individual level due to intermittent and irregular data collection. To synthesise activity-travel plans, we propose a novel model that combines mobile phone application data with travel survey data, addressing the limitations of each data type while leveraging their strengths. Our generative model simulates multiple average weekday activity schedules for over 263,000 individuals living in Sweden, approximately 2.6% of Sweden's population. These schedules include activity sequences, types, start/end times, and locations, incorporating daily plan variability. In this model, we propose a temporal-score approach to improve the state-of-the-art home and work location identification approaches among our designs to realistically synthesise activity-travel plans. We assess the generative model's performance against an existing large-scale agent-based model of Sweden (SySMo) and a dummy model using only mobile application data. The generated activity-travel plans are comparable to the SySMo model's output and significantly surpass the dummy model's results, suggesting the proposed model's capability to generate reasonable daily activity-travel schedules. The proposed model is adaptable to other regions with similar travel surveys and emerging data sources, like call detail records, advancing the use of these data for activity-based models in a cost-effective, easily updated manner.

Keywords Mobile phone application data · Big data · Activity-based modelling · Daily activity-travel plans

Introduction

Understanding individuals' mobility patterns is crucial for effective policy planning and service provision in transportation, healthcare, and urban development. In recent decades, activity-based travel demand modelling has become an important part of human travel behaviour modelling studies (Hilgert et al. 2017). This approach offers a better portrayal of human behavioural foundations influencing travel decisions compared to traditional population models. Travel surveys and diaries are commonly used as the principal data sources for these models, providing complete activity-travel plans and rich socio-demographic data (Huang et al. 2019). However, this data often comes with challenges, e.g., substantial implementation costs (Jiang et al. 2017), under-reporting of trips (Bricka and Bhat 2006), lower response rates among certain demographic groups (Ogle et al. 2005), and small sample size.

Mobile phone application data is a relatively new big data source and particularly valuable in transportation studies due to the precise location information enabled by the GPS functionality of the smartphone (Lee et al. 2016; Rodr et al. 2024). This data is collected by capturing the geographical locations of phone users as they interact with various mobile applications. The data enables large-scale observations of millions of individuals at a high spatial resolution over long time periods; this capability can be challenging and costly to collect with other data collection methods. Despite its potential, mobile application data presents significant challenges like most emerging data sources, e.g., call detail records (CDRs). These challenges include sampling biases due to uneven representation of different demographics and data sparsity caused by users not interacting with their phones, making it difficult to directly extract individuals' complete activity-travel patterns.

Although a few studies (Alexander et al. 2015; Jiang et al. 2017; Grujić et al. 2022) have shown the benefits of similar big datasets, e.g., CDRs for exploring individuals' daily mobility patterns, the use of mobile phone application data in activity-based models is under-explored. Moreover, these studies often overlook significant challenges, such as data sparsity, which limits the production of representative activity-travel patterns. At last, a notable gap is the lack of a generative model transferable to other similar big data sources.

Mobile phone data vary widely in their temporal resolution, spatial accuracy, and completeness, depending on the data source and collection method. This study focuses on one of the sparsest types, highly intermittent GPS records with substantial missingness, similar in limitation to CDRs (Fu et al. 2025). While other mobile phone signal data, such as signalling or app-based logs, may provide higher frequency and more continuous mobility trajectories, they still face challenges related to data gaps, particularly in specific environments like indoor areas, tunnels, or rural regions (Thierry et al. 2024). The model proposed in this study is explicitly designed to be robust under such sparse conditions, enabling it to extract meaningful mobility patterns even when observations are limited. Importantly, this robustness makes the model equally, if not more, applicable to richer datasets with lower levels of missingness. Across all mobile data types, handling sparsity remains a critical concern due to inherent technical limitations and contextual influences, reinforcing the relevance and flexibility of our modelling approach.

Our study introduces a generative model that synthesises activity-travel plans by combining mobile phone data with travel survey data (see details regarding the input data in Appendix section “Data”). This novel approach addresses the limitations inherent to these data sources, e.g., the sparsity of emerging big data and the small sample size of traditional

travel surveys, while leveraging their strengths, e.g., extensive observation periods of big data and complete one-day activity-travel plans of surveys. The proposed model simulates multiple average weekday activity schedules for over 263,000 individuals, approximately 2.6% of the population in Sweden. These plans are characterised by activity sequences, types, start/end times, and locations, incorporating daily variability in specific plan attributes. We also propose a temporal-score approach to improve the state-of-the-art home and work location identification approaches by considering the variations in activity engagement at different observation times.

To assess the generative model's performance, we compare it against the SySMo model (Tozluoğlu et al. 2023b), a large-scale agent-based model and a dummy model using only mobile application data (see details in section “[Dummy model](#)”). While the generative models' results are comparable to the SySMo model, they significantly surpass the dummy model's results. These comparisons highlight our approach's capability to generate realistic activity-travel schedules. The code for the generative models is available at <https://github.com/tozluoglu/MAD4AG>, and the methodology is adaptable to other geographical contexts by leveraging similar big data and survey data.

The remainder of this paper is structured as follows: Section “[Related work](#)” reviews the related literature. Section “[Model design](#)” describes the generative model for synthesising individuals' activity-travel plans. We introduce the model evaluation methods in section “[Model evaluation methods](#)”. Section “[Results](#)” presents the generative model's results and compares them with other models. In section “[Discussion and concluding remarks](#)”, we summarise our key findings, discuss the limitations of the generative model and provide suggestions for future work.

Related work

The rise of big data sources, e.g., location-based social networks (LBSNs), smart-card transactions, and mobile phone data, offers a promising alternative to develop activity-based models, traditionally reliant on travel surveys. These data provide a vast population and geographic coverage with superior spatial and temporal resolution (Lee et al. 2016). However, the characteristics of mobile phone data vary considerably depending on the collection method, which also influences the choice of appropriate methodologies. For example, mobile signalling data provide high-frequency, continuous mobility trajectories, while CDRs or application-based datasets often only capture partial location data. Moreover, these data sets may not fully represent the general population. Therefore, synthesising activity plans from sparse big data often faces significant challenges: (1) incomplete records of individuals' activity travel schedules, and (2) population biases.

Several studies have utilised sparse mobile phone data to generate individuals' activity-travel patterns, incorporating temporal, spatial, and semantic information about activities (Fu et al. 2025). For example, Jiang et al. (2017) extracted individuals' activity schedules, including activity types, location and sequences, from CDR data but did not model the temporal patterns of activity participation. Grujić et al. (2022) revealed car drivers' travel patterns combining mobile data with traffic counts, but this approach falls short in extracting temporal and semantic details about activities. Bassolas et al. (2019) generated activity-travel patterns from sparse mobile data by extracting activity sequences directly from the

data and assigning start and end times through random draws. Similarly, Widhalm et al. (2015) combined mobile data with land use information to identify activity locations, but inferred temporal patterns directly from the sparse mobile data.

To address the sparsity and capture the temporal patterns, Alexander et al. (2015) employed a travel survey as an auxiliary data source and identified individuals' daily schedules from CDR data. The study captures the temporal patterns by randomly sampling the trips' departure times based on the travel survey trip departure time distribution. However, the activity participation was inferred from the incomplete mobile phone records.

Although these studies demonstrate the potential of sparse mobile data for generating activity-travel patterns, their reliance on intermittent observations may lead to incomplete activity chains and missing trips due to periods when mobile users are not actively interacting with their devices (Ficek and Kencl 2012; Hoteit et al. 2014). Consequently, these studies may struggle to reconstruct the whole temporal structure of daily activity patterns.

A foundational step in activity-based models is the identification of primary activities, e.g. home, work, and school, due to their influence on individuals' daily plans (Arentze and Timmermans 2000; Bowman and Ben-Akiva 2001). Despite the spatial and temporal sparsity of mobile phone data, its large volume and extended observation period enable the inference of primary activity locations, as individuals tend to revisit these places regularly. Existing studies mostly rely on temporal rules to identify locations, assuming individuals will be in specific places during predefined hours. For instance, Chen et al. (2014) identify individuals' home and work activity locations by analysing daytime and nighttime periods, achieving success rates of 70% and 65% for home and workplace inference within 100 m compared with ground truth data. Sadeghinassr et al. (2019) detect home locations during the nighttime from 8:00 pm to 4:59 am. Additionally, Ahas et al. (2010), Kung et al. (2014), and Yu et al. (2020) infer individuals' home and work locations by applying distinct time frames using CDR data, providing essential information for evaluating commuting dynamics. While these assumptions are often valid, exceptions exist, and individuals' patterns of being at home or work can vary.

Among the emerging mobile data sources on human mobility, mobile phone application data is relatively new and has mostly been used in population-based modelling. This data shares characteristics similar to CDR data, e.g., collection methods through individuals' interactions with their mobile phones and data sparsity at the individual level due to intermittent data collection. Researchers have utilised mobile phone application data to generate OD matrices, illustrating trip distributions between locations (Zhang et al. 2019; Yu et al. 2020). Zhang et al. (2020) and Yu et al. (2020) used trajectories from mobile phone GPS datasets to identify individuals' travel modes. Woodard et al. (2017) predicted travel time variability at the trip and the road network link levels. Despite these studies, the application of mobile phone application data in activity-based models remains under-explored. Given its spatial precision compared to other big data sets, mobile phone data has the potential to accurately deduce activity locations, thereby enhancing the realism of activity-based models.

The existing literature highlights the benefits of emerging data sources in transportation studies. Despite the progress in leveraging these data, considerable gaps remain in generating realistic activity-travel plans. A significant challenge lies in addressing the sparsity of big data. Moreover, integrating mobile phone application data into activity-based models holds potential due to its spatial precision, large population coverage, and cost-effective-

ness, while it remains under-explored. Additionally, there is a need for more rigorous methods to identify activity locations, moving beyond evaluations based on temporal rules and visit frequencies. These enhancements would improve activity-based models' realism and advance our understanding of human travel behaviour.

Model design

We propose a generative model that simulates multiple average weekday activity-travel patterns by combining massive mobile phone GPS records with the national travel survey data. Appendix section “[Data](#)” introduces the input data, including anonymised mobile data and the travel survey. The model generates individuals who perform activities to fulfil various personal and social needs. The generated activities inform the overall travel demand of the population (Kitamura 1984), and can be applicable in simulations for infrastructure planning, public transportation optimisation, emergency response planning, environmental impact assessments, and economic analysis of transport policies.

The proposed generative model includes five modules (Fig. 1). First, we process mobile phone GPS records to infer key socio-demographic attributes, such as home location and employment status, thereby facilitating the synthesis of activity-travel plans. Second, we identify the locations of individuals' primary activities, i.e., home and work. We assume each individual has at least one home activity on an average day, and individuals with an inferred work location are classified as employees. Third, we design individual weights to have a representative sample of the Swedish population using inferred information. Fourth, we start searching for twin travellers in the travel survey by calculating the matching probabilities between people in the mobile phone application data and the travel survey participants. The matching probability is defined based on their demographic, activity patterns, and trip attributes. For each individual in the mobile phone dataset, we sample one twin traveller from the identified travel survey candidates. Finally, we generate activity-travel schedules for average weekdays using the selected twin travellers' daily activity sequence and individuals' activity locations in the mobile phone dataset. The proposed model is capable of synthesising different activity-travel plans for each agent, producing daily variations in individuals' activity-travel patterns. By iteratively finding twin travellers and activity locations, the model generates multiple average weekday simulations for each individual in the mobile phone dataset.

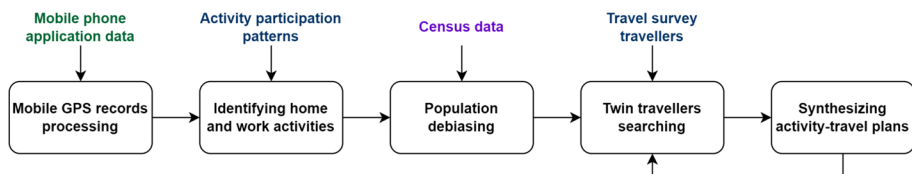


Fig. 1 Generative model for synthesising activity-travel plans. Travel survey inputs are marked with blue. Census data refer to the statistics related to Demographic statistics areas (DeSO zones), specified in Appendix section “[Demographic statistics areas \(DeSO zones\)](#)”

Mobile phone data processing

Through the GPS records from 322,919 mobile phone devices in Sweden, we first detect stops, which are defined as the geolocations where individual devices have spent more than 15 min. Based on the detected stops, we further identify activity records and their clusters. To improve the data quality for synthesising activity-travel plans, we also implement a few data filters, e.g., individuals with very few or outside Sweden records, etc.

Stop detection

To generate synthetic activity plans, it's essential to identify when and where individuals are engaged in various activities. This process begins by detecting "stops" from individual geolocation trajectories, which consist of user ID, latitude, longitude, and time information. A stop is where an individual remains for at least 15 min, characterised by latitude, longitude, start, and end times. These stop observations form the basis for creating activity clusters, identifying primary activity locations, and synthesising activity-travel plans in the proposed model.

Various algorithms can detect stops, as outlined in the literature (Ester et al. 1996; Aslak and Alessandretti 2020). For this study, we've chosen the Infostop algorithm (Aslak and Alessandretti 2020), known for its robustness against measurement noise, scalability with large datasets, and ability to process multi-user data concurrently. Details about the chosen algorithm and its parameter selection are discussed in Appendix section "Stop detection" (Liao et al. 2024). The stop locations are extracted over seven months in 2019 in Sweden.

Identifying activity records

Creating activity clusters Activity space is widely used to describe the geographical area where individuals conduct their daily activities within a time frame (Golledge 1997; Schönfelder and Axhausen 2003; Miller 2004). It encompasses the locations and spaces where people live, work, study, socialise, and engage in various other activities. After pinpointing the geolocations where individuals spend significant time, we form activity clusters to represent their daily activity spaces. This approach facilitates the creation of more representative daily activity-travel plans with less noise.

The geolocation data from mobile applications reveal distinct activity-travel patterns for individuals, e.g., one area with many activity locations in an urban area and one around a remote summer place. Contrary to earlier studies suggesting scale-free human movements (Brockmann et al. 2006; Gonzalez et al. 2008; Song et al. 2010), recent research shows that human mobility has definable scales, with individuals moving within and between different spaces confined by these scales (Alessandretti et al. 2020). Therefore, by examining individuals' activities and travel patterns, we identify activity clusters that reflect their daily mobility scales, excluding rare events such as residential mobility from one area to another.

To determine individuals' activity clusters, we apply the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, which has been widely used in the literature (Sadilek et al. 2012; Stalla-Bourdillon and Knight 2016; Wojtusiak and Nia 2021). The DBSCAN algorithm clusters points based on proximity, requiring a minimum number of points (*MinPts*) within a specified radius (ϵ) (Ester et al. 1996). We set the ϵ parameter

to 200 km based on the 99th percentile of travel distances from the travel survey and the *MinPts* parameter to 2 to ensure the detection of at least one additional activity alongside the home activity within each cluster. This process allows us to establish spatial scales where we infer typical weekday activity patterns, treating activity clusters that are distant from each other separately. For example, an individual residing in one location for three months and then another for three months would have two activity spaces, which can be identified and analyzed separately by the model.

Building-level activity records After identifying activity clusters, we group stops of each device in each activity cluster to determine where individuals engage in daily activities, e.g., home, work, and others, thereby generating activity records. Each stop offers a glimpse or segment of individuals' activities at a given location. To identify activity records, the DBSCAN algorithm is applied to the stops' GPS coordinates, with the ε parameter set to 100 m, reflecting the spatial resolution observed in the data. This method groups the stops based on proximity.

However, one activity may end up in different locations for two reasons: noise from GPS records and movement within the activity location. First, factors such as GPS receiver type, signal quality, and environmental conditions cause variations in spatial resolution of stops, ranging from centimetres to hundreds of meters (Thierry et al. 2024). Second, the small movements of individuals while engaged in activities further disperse the stop locations, e.g., moving between offices at work.

To address the scattering of identified activity records around the buildings, we generate building-level activity records that link the activity records with buildings for use in further analysis. We derive representative GPS coordinates for activity locations by calculating the mean values of the coordinates of the stops within each group. Subsequently, these calculated coordinates are snapped with the nearest building to represent the recorded activities' geolocations.

Filtering data

We select individuals and their activity locations for further analysis based on specific criteria to guarantee the data quality. Activities that exceed 12 h are excluded, as are those outside of Sweden. We also exclude individuals with fewer than seven active days, where an active day is defined as having at least one recorded activity. Activities collected on weekends and holidays are not considered. Lastly, individuals with fewer than two unique activity locations within their activity clusters are removed from the analysis. These criteria ensure a focused dataset for our study.

After processing, the data contains 287,217 individuals with 12,740,559 activity records distributed across 312,901 identified activity clusters. Table 1 shows statistics about individuals and their activity records.

Table 1 Descriptive statistics of activity records. An active day is a day with at least one recorded activity

Attribute	min	25%	50%	75%	max
No. of activities	2	10	21	48	1129
No. of unique activity locations	2	3	4	10	374
Median activity duration (min)	16	180	182	192	717

Temporal-score approach for inferring primary activities

We infer primary activity locations for individuals in the mobile phone application data. Primary activities (or anchor activities) are critical activities, e.g., work and school, whose locations and timings are determined independently of other activities (Arentze and Timmermans 2000; Bowman and Ben-Akiva 2001). These activities play a role for individuals to plan their secondary activities, including shopping, leisure, and social visits. In this study, we define two categories under primary activities: home and work activities. The work activity category also includes educational activities. Existing methods often rely on temporal rules to identify residential locations (Yabe et al. 2022; Vanhoof et al. 2020; Jiang et al. 2017). These rules include that individuals are mostly at home during nighttime and that the most frequently visited place is the individual's home. Although these assumptions hold in many cases, some exceptions exist, and the patterns of being home and working may vary among individuals.

To infer home locations, we propose a temporal score method that accounts for the varying likelihood of presence at an activity location across different hours of the day. We infer the primary activities based on the activity cluster where they are situated. We first calculate the number of visits to each activity location during nighttime over the observation period. The definition of nighttime varies across studies, e.g., 8:00 pm to 4:59 am (Sadeghinasr et al. 2019) and 7:00 pm to 7:59 am (Vanhoof et al. 2020). In this study, we set a temporal rule for nighttime based on the travel survey data, defining it as the hours in a day when the rate of home activity participation surpasses 0.8 of the population, i.e., the hours where 80% of the travel survey participants are home. Accordingly, the nighttime starts at 6:00 pm and ends at 7:59 am the following morning. Each visit to an activity location during this time frame is counted as one per day, regardless of whether the location is visited one or more times. The activity locations being visited fewer than three times during the data-informed nighttime are excluded from further analysis.

Next, we deal with the remaining activity locations being visited at least three times during the pre-defined nighttime. In this step, we calculate a score, considering the variations at the observation time rather than equally treating the hours within the nighttime. For example, according to the travel survey data, more individuals are home at 3:00 am than at 7:00 am. Therefore, an activity location visited at 3:00 am is more likely to be home than at 7:00 am. To calculate such scores, we design weights for each nighttime hour between 6:00 pm and 7:59 am, derived from the ratio of hourly home activity participation to overall activity participation in the travel survey (Fig. 9). Accordingly, we weigh the duration of an individual's stay at each activity location during the night hours and calculate its score (detailed in Appendix section "[Temporal-score approach for inferring primary activities](#)").

We identify home activity locations with the highest score above a score threshold and being at least visited three times during the data-informed nighttime. The score threshold is set to 10 by evaluating the inferred residential location's building type. The threshold is chosen to balance a reasonable share of home activity locations at residential buildings and sufficient data. We identify the residential locations of over 263,000 individuals in the mobile phone dataset. Individuals whose home locations could not be inferred are excluded from further analysis based on our assumption that every individual participates in a home activity.

To identify work locations, we adopt a similar approach to the residential location inference. After excluding the inferred home locations from the activity locations, we evaluate typical working hours, specifically from 8:00 am to 5:59 pm, according to the travel survey. We then calculate a score for each remaining activity location (detailed in Appendix section “[Temporal-score approach for inferring primary activities](#)”). To ensure the quality of workplace location inference, we remove activity locations with a score below 30. We keep the activity location with the highest score among all evaluated locations as a workplace. We identify the workplaces of approximately 67,000 people, accounting for 26% of the population whose home locations are identified.

Population debiasing

To ensure a representative sample of the Swedish population, we compute a weight for each individual in the mobile phone data. These weights address the inherent sampling bias arising from the spatial distribution of the individuals in the mobile phone dataset across Sweden. To calculate weights, we first associate individuals in the mobile phone dataset with Demographic statistics areas (DeSO zones, specified in Appendix section “[Demographic statistics areas \(DeSO zones\)](#)”), using their inferred home locations. We then apply these zones’ population and employee sizes in the following weight design.

The weight calculation has two stages: First, we apply inverse probability weighting (IPW) to assign initial weights based on the ratio of the calculated population size to the adult population size in the DeSO zones. Second, we adjust these weights to match the distribution of employees across different zones. We use the Iterative Proportional Fitting (IPF) technique that iteratively adjusts the elements of a matrix to ensure that the sums across each row and column align with predetermined totals (Deming and Stephan 1940). In each DeSO zone, we scale the initial weights of employees by the actual employee size in the statistics while maintaining the overall population size (Equation 1). For example, if employees in a DeSO zone are expected to constitute 50% of the population, but only 40% are observed in our dataset, IPF addresses this discrepancy by increasing the weights of employees and reducing the weights of non-employees with a multiplier reflecting the accurate employment rate, while maintaining the population size in each DeSO.

$$\begin{aligned} W'_{p,e} &\leftarrow \left(\frac{N(z,e)}{\sum_{p \in z,e} W_p} \right) \cdot W_{p,e} \\ \forall (p \in z) : \quad W'_p &\leftarrow \left(\frac{N(z)}{\sum_{p \in z} W_p} \right) \cdot W'_p \end{aligned} \quad (1)$$

where p denotes individuals and W_p their initial weights, e is employees. $N(z, e)$ represents the desired number of employees in zone z_d , $\sum_{p \in z,e} W_p$ is the sum of initial weights for employees in that zone, and $\sum_{p \in z} W_p$ is the sum of weights in the zone. W'_p reflects the adjusted value after the IPF process.

To prevent extreme weights, we apply the weight-trimming technique, where any weight above the threshold (W_0) is set to W_0 (Van de Kerckhove et al. 2014). Equation 2 calculates this threshold.

$$W_0 = 3.5 \times \sqrt{1 + CV^2(W_p) \times \text{Med}(W_p)} \quad (2)$$

where W_p denotes the individuals' weights, CV is the coefficient of variance, and Med is the median value of the weights.

Searching twin travellers

The model synthesises activity-travel plans from abundant but incomplete activity records from mobile phone data. To do so, we first synthesise activity sequences for individuals in the mobile data, considering their observed mobility patterns and their similarity to the travel survey participants who have complete activity records. We adopt a statistical matching approach using the travel survey data. Statistical matching with some variations is often applied in activity-based modelling to define daily activity patterns for synthesised individuals based on their similarity to survey participants (Hörl and Balac 2021; Namazi-Rad et al. 2017).

The statistical matching approach connects each individual in the mobile phone data with one travel survey participant, defined as his/her "twin traveller." We first group individuals in the mobile phone data and the travel survey data based on key mobility attributes, e.g., employment status and average trip distance, as detailed in Table 5. For mobile users, these attributes are inferred from GPS traces, while for survey participants, they are derived directly from survey responses. We then calculate a matching likelihood probability between the groups. Using these probability results, we pair each individual in the mobile phone data with one twin traveller in the travel survey participants considering their shared mobility attributes.

Categorization of travellers

We group individuals in the mobile phone and travel survey data based on their demographic, activity, and trip attributes. The considered attributes are the individual's residential region (Svealand, Götaland, or Norrland), urban density of residence (high or low), employment status (yes or no), average trip distance between home and activities (long or short), and commuting distance (long or short). For the individuals in the mobile phone data, we use all the attributes. For the travel survey participants, however, we adopt a step-wise grouping approach due to the limited sample size. The detailed definitions of these variable levels are in Appendix section "[Searching twin travellers](#)".

After this step, all individuals in the mobile dataset are assigned a group (c_i for individual i). For example, individuals with a high average trip and commuting distance residing in a low urban density area in Götaland.

Sampling twin travellers

Classifying mobile phone users and travel survey participants into the same category of travellers creates a common basis for the next step: identifying a twin traveller for each individual in the mobile phone dataset within each category. Instead of randomly selecting the twin traveller within the same category, we calculate matching likelihood probabilities

for each travel survey participant, considering the resulting distributions of commuting and activity patterns at the aggregated level.

We calculate the matching likelihood probabilities using the IPF technique. We start by assigning a uniform probability to each individual within a category. This initial probability is inversely proportional to the total number of individuals in that category. We then adjust these probabilities using multipliers that reflect the distributions of commuting and activity sequences observed in the survey data. We iteratively scale the probabilities to match the cumulative probability distribution with the respective marginal distributions. After each adjustment, we also normalise the probabilities for each individual to ensure that the sum of probabilities within each category equals the total number of individuals in that category. This iterative fitting process continues until the adjusted probabilities converge, indicated by negligible changes between subsequent steps. Equations 3–5 show the calculation of the matching likelihood probabilities.

$$p_{i,k,s}^{(t+1/3)} = p_{i,k,s}^{(t)} \frac{D_k}{\sum_{i,s} p_{i,k,s}^{(t)}} \quad (3)$$

$$p_{i,k,s}^{(t+2/3)} = p_{i,k,s}^{(t+1/3)} \frac{S_s}{\sum_{i,k} p_{i,k,s}^{(t+1/3)}} \quad (4)$$

$$p_{i,k,s}^{(t+1)} = \frac{p_{i,k,s}^{(t+2/3)}}{\sum_{i,k,s} p_{i,k,s}^{(t+2/3)}} \quad (5)$$

where $p_{i,k,s}$ is the probability assigned to each individual i , commuting type k , and activity sequence type s in category c_i , D is the vector of marginal probabilities for commuting, and S marginal probabilities for activity sequences. After initializing the individual weight $p_{i,k,s}^{(0)}$ for individual i , we iteratively adjust the weight $p_{i,k,s}$ to fit the marginal probabilities D and S . In each iteration t , we perform the row adjustment for commuting probabilities (Equation 3), the column adjustment for sequence probabilities (Equation 4), and normalisation (Equation 5). This iterative fitting process continues until the adjusted probabilities converge, indicated by negligible changes between subsequent steps.

To find twin travellers for individuals in the mobile data, we first identify candidate individuals within each group for the travel survey participants. Using the calculated matching probabilities, we randomly sample a set of survey participants to equal the number of individuals in the mobile dataset for each group. Secondly, we rank individuals from both groups in ascending order by their average trip distance and match them one-to-one. During the matching process, we ensure that individuals without identified secondary activity locations are not matched with survey participants who have such activities in their sequence. Ultimately, we assign the individuals in the mobile phone dataset with the activity types, start and end times, and sequences of their twin travellers in the travel survey.

Synthesising activity-travel plans

After identifying a twin traveller for each individual in the mobile phone dataset, we have activity schedules assigned. In this section, we show how we continue assigning primary and secondary activity locations to the corresponding activity schedules.

Location assignment of other activities

Other activities (secondary activities) are all activities apart from primary activities in a daily activity schedule, e.g., shopping, leisure, etc. The mobile phone dataset, covering seven months of observation, features abundant other activity locations for individuals, providing valuable variations in travel demand induced by secondary activities. On average, each individual has 6.6 other activities, with approximately 18% individuals having more than ten other activity locations. We thus need to determine which other activity locations are visited on an average weekday in the synthesised activity-travel plans.

For all other activities in the individual's schedule, we select the corresponding locations by considering their distances to their home location and the number of other activities in the individual's activity sequence. We define the probability of selecting one other location as below:

$$P_o \propto \frac{f_o}{\log(d_{o,h} + 1)} \quad (6)$$

where P_o represents the probability of selecting one other activity location, f_o the frequency of visits to this location, h the individual's home location, and $d_{o,h}$ the distance between them.

The individuals' probability of visiting each other activity location is normalised by all other location records so that the probability of visiting all adds up to 1. Using the calculated probabilities, we randomly sample other activity locations for all other activities in the individual's activity sequence.

Completing activity-travel plans

Now, we assign an activity sequence for each individual in the mobile phone dataset, including activity types, start/end times, and where these other activities occur, if any. We have also previously inferred the primary activity locations, i.e., home and work, from the steps in section "[Temporal-score approach for inferring primary activities](#)". We first place these primary activity locations into the individuals' activity schedules. We then allocate the selected other activity locations to the corresponding activities in the schedule. The other activities are arranged chronologically and positioned based on proximity to the most recently placed primary activity. In other words, the other activity location is assigned if it is the closest to the preceding primary activity location.

The proposed generative model is capable of generating multiple average weekday activity-travel plans for one individual in the mobile phone dataset. One can repeat the steps of searching for twin travellers (section "[Searching twin travellers](#)") and synthesising activity-travel plans (section "[Synthesising activity-travel plans](#)") to get a variety of plans for indi-

viduals. At the individual level, variations occur across all elements of activity schedules, i.e., activity sequences, type, start/end times, and other activity locations, while maintaining the residential and workplace locations. Despite these variations at the individual level, the activity patterns at the population level remain robust by design.

Model evaluation methods

We assess the performance of the generative model by comparing its activity-travel schedules with the travel survey data and other similar activity-based models. The performance is evaluated by reflecting the key aspects of travel demand, e.g., daily activity sequence distribution, temporal activity engagement patterns, and trip distance distribution. We include comparisons with a dummy model, which adopts a simplified methodology using only mobile phone datasets to generate average weekday activity-travel schedules, and the Synthetic Sweden Mobility (SySMo) model, a large-scale agent-based transportation model of the Swedish population. These comparisons provide insights into the generative model's ability to integrate mobile data with survey data to generate realistic simulations.

Dummy model

To serve as a baseline for comparisons with the generative model, we present a dummy model that adopts a simpler methodology and uses only the mobile phone dataset. The model generates average weekday activity-travel schedules characterised by activities' type, start and end times, duration, sequence, and location.

The model includes three steps: mobile phone data processing, inferring activity locations, and generating activity schedules. In the mobile phone data processing step, we adopt the approach used in our generative model, detecting stop points where the individual remains stationary for significant periods and identifying activity records from these stops. In the second step, the model infers primary and secondary activity locations categorised using the state-of-the-art approaches instead of the temporal-score approach in this study (Section "[Temporal-score approach for inferring primary activities](#)"). The primary activities are home and work, and the secondary activities are all remaining activities, e.g., shopping and restaurant visits. The model determines primary activity locations by counting the number of visits to each location during designated hours: nighttime for homes (6 pm to 7:59 am) and off-nighttime hours (8:00 am to 5:59 pm) for workplaces. The location with the highest visit count during these hours is assigned as the respective primary activity location. The remaining abundant activity locations visited during the observation period are categorised as secondary activity locations.

The final step includes generating individual activity-travel schedules from the mobile phone dataset. For each individual, the model calculates the hourly frequency of visits to each activity location over the observation period. The model assigns the most frequented location during each specific hour of the day as the activity for that time. Using previously inferred activity types, the model constructs daily activity schedules.

The dummy model relies solely on mobile phone geolocation data and generates activity-travel schedules without considering the sparsity of the dataset as well as the population biases. It provides a data-driven representation of the spatio-temporal mobility patterns of

its covered mobile devices. The methodological differences between the proposed generative model and this dummy model are summarised in Table 2.

Synthetic Sweden Mobility (SySMo) model

The Synthetic Sweden Mobility (SySMo) model (Tozluoğlu et al. 2023b) is a large-scale agent-based transportation model. It simulates the Swedish population and their transport behaviours on an average weekday. The model generates a synthetic replica of over 10 million individuals with socio-demographic attributes, e.g., age, gender, civil status, etc. These attributes are the basis for travel demand generation. Using an activity-based approach for demand generation, the model assigns daily activity-travel schedules, including the type of activity, start-end time, duration, sequence, the location of each activity, and the travel mode between activities to each individual in the population.

The model's results are evaluated against official statistics from Statistics Sweden and Trafikanalys and are publicly available in a data repository (Tozluoğlu et al. 2023a). The model outputs, which include agents' socio-demographic information and mobility patterns, have served as a base for other research, e.g., Liao et al. (2023), Tozluoğlu et al. (2024).

Performance metrics

To evaluate the performance of the generative model, we conduct a series of comparative analyses against the dummy and the Synthetic Sweden Mobility (SySMo) models. These analyses focus on three key aspects of human mobility: daily activity sequences, temporal activity participation patterns, and trip distances. We choose these comparisons to provide

Table 2 Model methodology comparison between dummy and proposed generative model

Model component	Dummy model	Proposed model
Primary activity inference	Infers primary activities based on visitation frequency	Considers stay durations using the temporal-score approach informed by survey data
Secondary activity inference	Uses only secondary activity locations with the highest hourly visitation frequencies	Incorporates all observed secondary activity locations in the data, considering their visitation frequency and distance from primary activities
Activity sequence generation	Derives activity sequence distribution from the input data	Mirrors activity sequence distribution in the travel survey data
Activity schedule generation	Generates activity sequences and start/end times based on hourly visitation frequency, relying on abundant but incomplete activity records	Copies activity sequences and start/end times from twin travellers with complete activity records
Simulation	Generates a single average day activity-travel patterns	Generates multiple average day activity-travel patterns using its probabilistic methodology

a comprehensive assessment of the generative model’s ability to represent the mobility behaviour of the Swedish population.

To quantify the similarity of distributions across different models, we use the Jensen–Shannon (*JS*) distance. The *JS* distance is a symmetrized and smoothed variation of the Kullback–Leibler (*KL*) divergence (Österreicher and Vajda 2003). To compute the *JS* distance, we first calculate the *KL* divergence for each pair of distributions. The *JS* distance is then derived as the square root of the average of the *KL* divergences between each distribution and the mean distribution, *M*, of the pair (*p* and *q* distributions). This method results in a distance range from 0 to 1, where 0 indicates identical, and 1 is a different distribution. We calculate the *KL* divergence and the *JS* distance as below:

$$KL(p, q) = \begin{cases} p \log(p/q) - p + q & p > 0, q > 0 \\ q & p = 0, q \geq 0 \\ \infty & otherwise \end{cases} \quad (7)$$

$$JS(p, q) = \sqrt{\frac{KL(p, M) + KL(q, M)}{2}}$$

Results

This section presents the results of the generative model that synthesises activity-travel schedules by fusing the mobile phone application and the travel survey data. We also assess the model’s performance by comparing its outputs with various sources, including grid-level population statistics, the SySMo and dummy models, and travel survey data that served as one input for the generative model.

First, we present the primary activity inference results and compare them with the actual population size. Next, we examine the distribution of activity sequences and temporal activity participation across the models and the travel survey. Finally, we compare trip distances and commuting distances across the different models. This comparison helps to highlight the model’s strengths and potential areas for improvement in synthesising activity-travel patterns using big geolocation data sources.

Primary activity locations

The generative model identifies the individuals’ primary activity locations, i.e., home and work locations, from the mobile phone application data. These activities are crucial for constructing activity schedules, and their spatial distribution reflects the demographic representativeness of the data.

We analyse the spatial distribution of primary activities to give a more comprehensive perspective on primary activity inference. Figure 2a and b illustrate the distribution of identified home locations compared to the actual population sizes. The generative model identifies home locations in each DeSO zone. Figure 2a compares the identified home locations with the population size at the DeSO zones. In most zones, the share of the number of home locations in the total population is between 1% and 5%. We also calculate a Spearman cor-

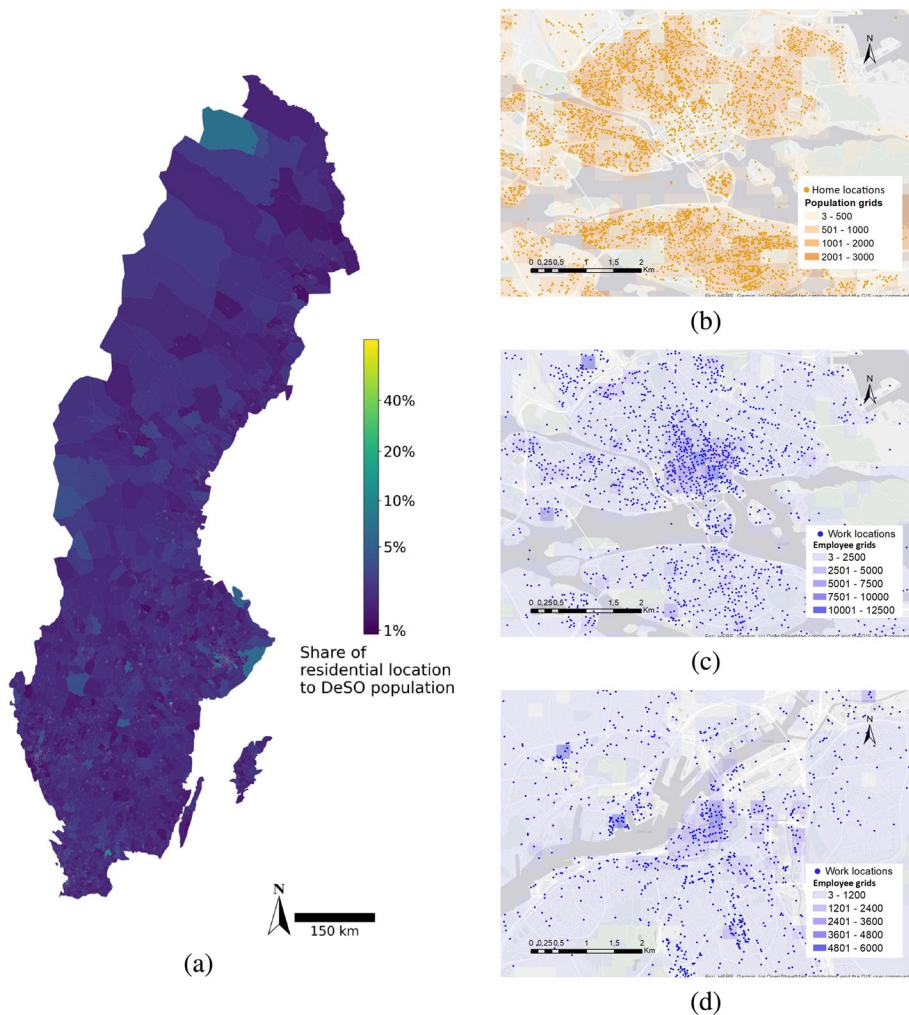


Fig. 2 Identified home and work locations vs. census statistics: **a** Share of the inferred home locations to the DeSO zone population (in %). **b** Comparison of the inferred home locations with the population size in 250 m grids (SCB) across central Stockholm. **c** Comparison of the inferred workplace locations with the number of employees in 250 m grids (SCB) across central Stockholm. **d** Comparison of the inferred workplace locations with the number of employees in 250 m grids (SCB) across central Gothenburg

relation coefficient of 0.49 ($p < 0.001$) between the number of identified homes and population size. This value indicates a moderate representation of the population's magnitude in each DeSO zone. Figure 2b further explores the home location distribution with population density at a 250 m grid level in central Stockholm, showing the alignment of home location inference with statistical data.

Further, we assess the identified workplace locations to evaluate the generative model's performance. Figure 2c and d present the spatial distribution of workplace locations compared to the number of employees at 250 m grids in central Stockholm and Gothenburg, respectively, the two most populated cities in Sweden. High employee density zones show

a dense concentration of identified work locations, affirming the model’s ability to capture employment hubs.

Daily activity sequences

Activity sequences show the order of activities during an average day, e.g., home-work-home (H-W-H). Figure 3 illustrates the distribution of the eight most common daily activity sequences, which account for over 75% of all sequences generated by the generative model, and compares it with the other models and the travel survey.

The generative model produces a similar activity sequence distribution to the Synthetic Sweden Mobility (SySMo) model and the travel survey data. However, it slightly overestimates the Home-Other-Home sequence. This discrepancy is likely due to the mobile phone data, which provides a more comprehensive view of secondary activities that can be under-reported in survey data (Lee et al. 2016).

In contrast, the dummy model does not fully capture the distribution of the sequences compared to the Synthetic Sweden Mobility (SySMo) model and the travel survey data. It has a low share of the Home-Work-Home sequence, one of the dominant activity sequences, and other less frequent sequences, e.g., Home-Work-Home-Other-Home. This discrepancy in infrequent sequences highlights the dummy model’s limitations in modelling complex activity patterns.

To quantify the similarity in distribution across the models, we use Jensen–Shannon (JS) distances (Endres and Schindelin 2003), ranging from 0 to 1, where 0 indicates identical distributions. The JS distance between the generative model and the SySMo model is 0.12. Similarly, the distance between the generative model and the travel survey is 0.07. Both

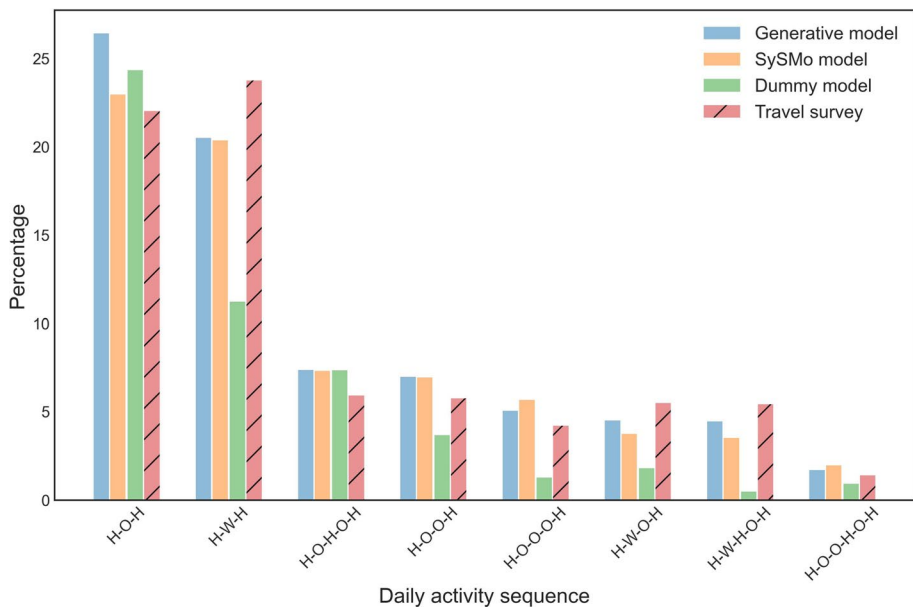


Fig. 3 Share of eight most frequent daily activity sequences. Home: H, Work: W, Other: O. These sequences comprise over 75% of all sequences by the generative model results

distances confirm close alignment in their sequence distributions. This contrasts with the dummy model's worse performance (JS distance= 0.41), indicating a lower level of alignment. This substantial difference suggests that the generative model's capability to synthesising activity sequences is notably superior compared to the dummy model. In other words, relying solely on the mobile phone geolocation data is insufficient to produce reasonable and useful activity sequences.

Temporal activity patterns

We assess hourly activity patterns to better understand temporal changes, which are vital for effective scheduling. Figure 4 shows the temporal distribution of activity participation by activity types from midnight to the following day, comparing results across different models and the travel survey. The generative model follows the home activity patterns compared to the SySMo model and the travel survey data during both night and daytime hours. It also has work and other activity engagement patterns similar to SySMo and the survey.

The dummy model captures the general pattern of hourly activity participation but underrepresents home activity participation during the night hours. Moreover, compared to the SySMo model and travel survey data, it overrepresents other daytime activities. These discrepancies underscore the dummy model's limitations and, conversely, demonstrate the generative model's ability to capture both nighttime and daytime activity patterns accurately.

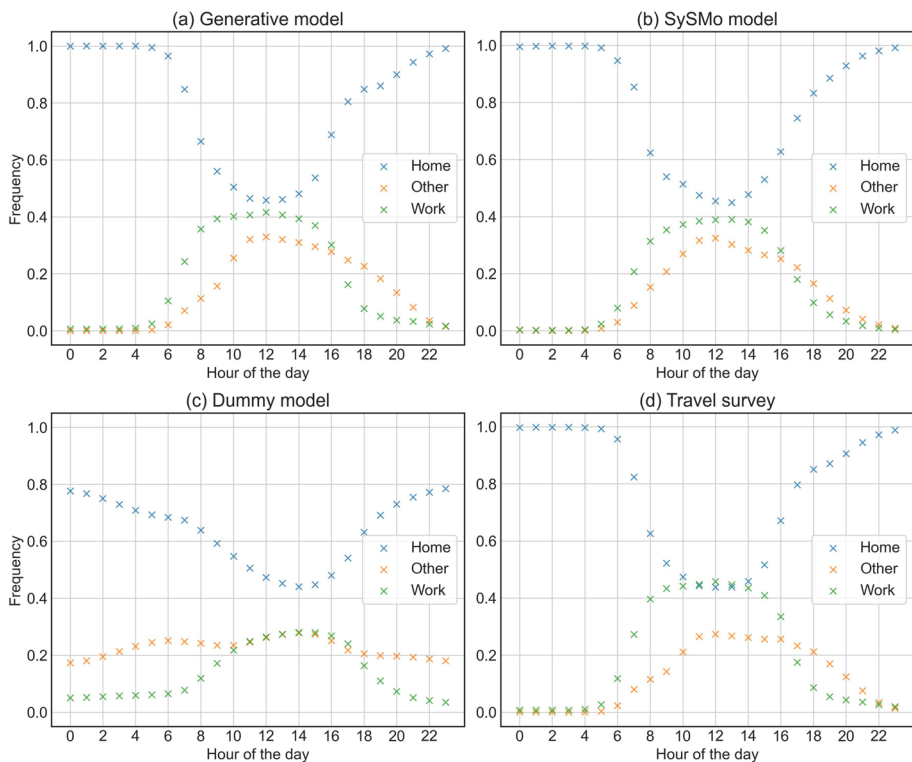


Fig. 4 Aggregated temporal activity engagement patterns by activity type. Frequency is the share of individuals being at a certain activity. **a** Generative model, **b** SySMo model, **c** Dummy model, **d** Travel survey

The spatio-temporal activity engagement patterns by type in central Stockholm (Fig. 5) corroborate with the above observations. Home activities are prevalent throughout the day in residential areas, increasing during night hours. Work activities peak during daytime hours, particularly in central business districts, e.g., Norrmalm and Östermalm. While the distribution of other activities is geographically similar to work activities, their occurrence extends into night hours, illustrating a distinct temporal variation. Figure 10 in the appendix shows the activities' geolocations by the time of day in central Stockholm.

Trip distances

We evaluate the generative model's travel patterns by comparing Euclidean trip distances between activity locations across different models and travel survey data. This analysis presents the model's ability to simulate daily trips.

The generative model captures the trip distribution patterns observed in the SySMo model and the travel survey. Figure 6 illustrates the trip distance distribution, and Table 3 compares the mean, median, and 90th percentile values of trip distances across the models. Most trips fall within the 0–20 km range across all models and the travel survey data. The generative model exhibits a higher frequency for short-distance trips of less than 4 km and tends to overestimate distances for trips exceeding 60 km. As a result, it reports higher mean and 90th percentile values than the travel survey and SySMo model, reflecting its inclusion of longer trips, which are often underreported in traditional travel surveys (Janzen et al. 2018). However, with a median trip distance of 3.7 km, the generative model is comparable to the other sources, indicating its ability to capture central trip distances despite the mobile data including a broader range of trips. The mean and median commuting trip distances of the generative model also align well with the SySMo model and the travel survey. In contrast, the dummy model consistently shows the highest values, largely diverging from the survey data and the SySMo model.

Robustness and variance

The generative model creates multiple average weekday simulations by iteratively finding twin travellers and activity locations for each individual in the mobile phone dataset. This produces variance in the resulting activity-travel plans while maintaining activity patterns at the aggregate level.

First, the model introduces variations in all components of the activity schedules, including activities' sequence, types, start/end times, and other activity locations while preserving the residential and workplace locations. Figure 7a illustrates the diversity in activity sequences and schedules over five simulation days. While more than half of the individuals have the same or only two different activity sequences throughout the simulation period, their activity sequences show variation because of the activities' start/end time and the location of other activities. This variation adds depth to the simulation by reflecting realistic daily dynamics, leveraging abundant geolocations for the other activities from mobile phone data.

Second, the model produces stable aggregated activity patterns despite the individual variance. Figure 7b compares aggregated activity engagement by activity types across five simulation days. Despite the observed variations at the individual level, there is minimal

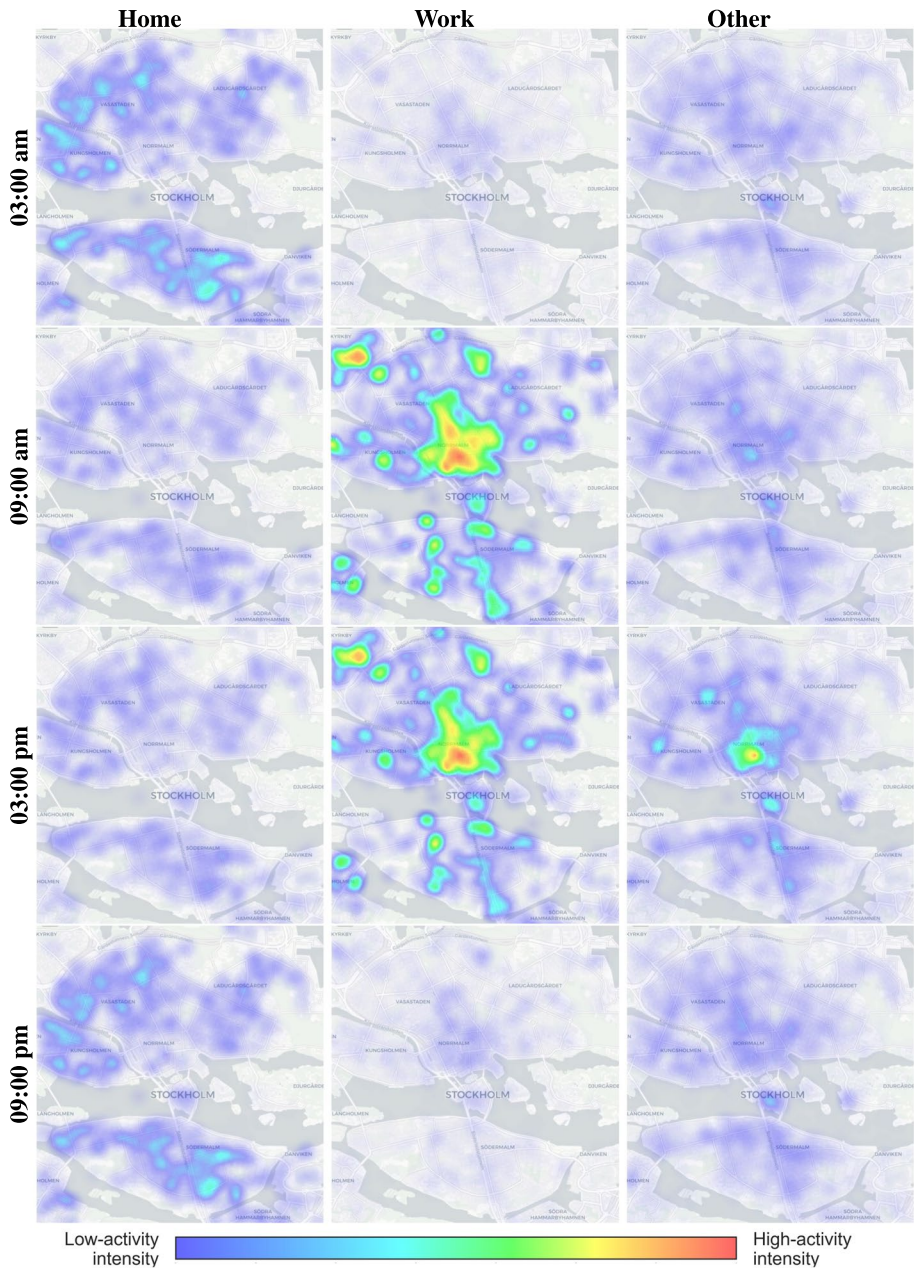


Fig. 5 Spatio-temporal distribution of activity participation by type in central Stockholm. The warm colours indicate hotspots of high activity intensity, while cold colours indicate relatively low activity intensity for a certain activity type

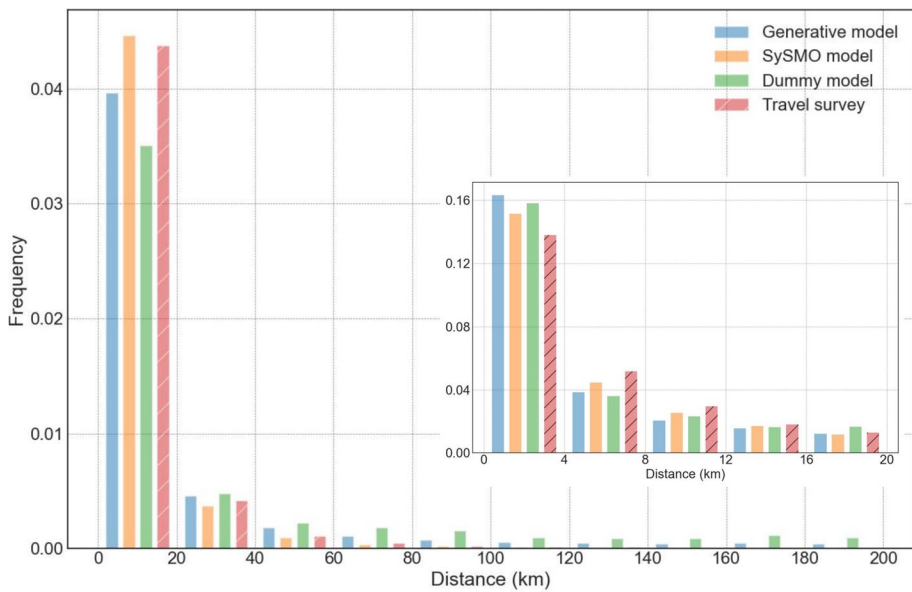


Fig. 6 Trip distance distribution by models (in the range of 0–200 km). The inset zooms into the 0–20 km range

Table 3 Trip distances comparison by models (in km)

Model	Trip type	Mean	Median	90th percentile
Generative model	Overall	21.5	3.7	56.3
SySMo model	Overall	9.8	3.4	21.3
Dummy model	Overall	35.3	6.5	124.6
Travel survey	Overall	12.5	4.3	24.1
Generative model	Commuting	18.3	5.3	39.7
SySMo model	Commuting	11.7	5.6	27.6
Dummy model	Commuting	32.9	8.4	104.4
Travel survey	Commuting	15.5	7.9	30.24

fluctuation in overall activity participation across the simulations, indicating consistent behaviour patterns among the simulated population.

Discussion and concluding remarks

This study introduces a generative model that synthesises individuals’ daily activity-travel patterns from extensive mobile phone GPS records. Mobile phone application data, a relatively new big data source, offers the potential for generating detailed activity-travel patterns, but its individual-level sparsity complicates creating realistic activity plans. Previous studies have highlighted the benefits of similar big data sources in activity-based modelling studies but have not adequately addressed data sparsity and population bias issues (Alexander et al. 2015; Jiang et al. 2017; Grujić et al. 2022). By combining mobile data with a travel

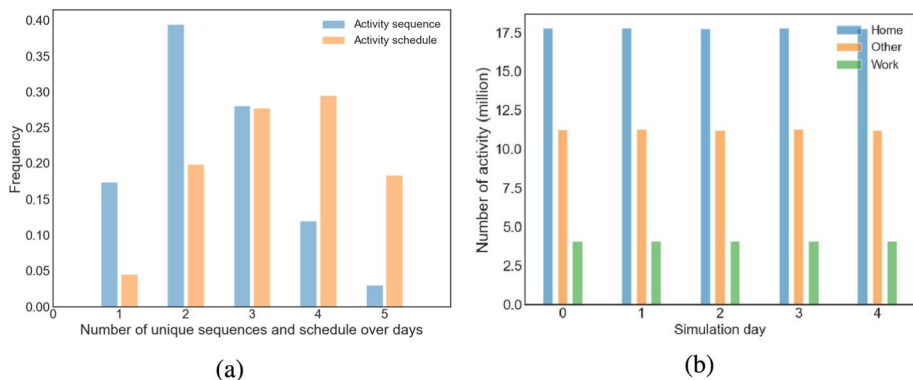


Fig. 7 Comparison of multiple average weekday simulations: **a** the number of unique activity sequences and schedules over five simulation days, **b** Aggregated activity engagement by activity types across five simulation days

survey, our study addresses mobile data's sparsity and biases and realistically simulates average weekday activity schedules for 263,000 individuals in Sweden.

To assess the generative model's performance, we compare its results with those from the underlying travel survey data, Synthetic Sweden Mobility (SySMo) model, and the dummy model that relies solely on mobile phone data. The generative model's activity pattern closely aligns with the underlying survey data and the SySMo model, substantially outperforming the dummy model. The generative model's activity sequence distribution achieves a Jensen–Shannon distance of 0.07 compared to the survey and 0.12 compared to the SySMo model, indicating high similarity. Additionally, the comparison of hourly activity participation during the day further confirms the close alignment of the generative model with the survey data and the SySMo model. Meanwhile, the dummy model deviates from the realistic pattern revealed by the other models. The generative model accurately captures trip distance distributions in travel patterns, with median trip distances aligning with the survey data and SySMo model. The dummy model shows the highest values, diverging from both the survey and SySMo models.

Furthermore, our study advances the state-of-the-art primary activity identification approaches by introducing a temporal-score approach. In anonymised mobile phone data, these locations are typically inferred using temporal rules (e.g., the studies by Chen et al. 2014; Sadeghinasr et al. 2019), often failing to account for individual differences in activity participation patterns. Our approach considers variations in activity engagement at different observation times. The inferred home and work locations align well with the grid-level population and employee statistics, demonstrating the model's capability to accurately identify primary activities.

While the generative model effectively generates activity-travel patterns from mobile data, it is also important to acknowledge the model's limitations and consider the broader context of activity-based models. The model captures key attributes of activity-travel plans, including activity type, start and end times, duration, sequence, and location, but does not include travel mode information between activities. Incorporating a mode choice component could further enhance the model's applicability. Previous research informs that the sparsity of mobile phone data makes it challenging to extract fine-grained mode informa-

tion (Huang et al. 2019), although recent studies have shown promise for mode inference using mobile data alone or in combination with survey data (Chin et al. 2019; Wischer et al. 2023). In addition, the generative model excludes short-duration activities. Our methodology defines activity stops as locations where individuals remain for more than 15 min, which systematically omits short activities such as meal pick-ups or drop-offs. Including these short stops could further improve the realism of generated activity-travel patterns. Both of these limitations present opportunities for future research.

One major challenge lies in the disaggregated evaluations of the activity travel schedules. Most models in the literature lack validation due to the absence of ground truth data. Despite this study's promising findings, this gap limits our ability to test the model's reliability and accuracy in a precise way. Furthermore, the combined data sources have different collection periods: survey data from 2011 to 2016 and mobile application data from June to December 2019. Future work should carefully align the data sources and explore any possible travel behaviour changes between years in Sweden.

Another limitation is that the model captures variability in travel behaviour by generating multiple plans, but only for an average weekday. This means that the model does not account for weekly variations of activity demand. Analysing activity patterns over extended periods, such as a week, could provide valuable insights into infrastructure measures and travel behaviour stability and variability (Moeckel et al. 2024). One way to address this limitation is to integrate week-long travel diaries, enabling more comprehensive model calibration. Future research aims to address this limitation by extending the model to weekly activity pattern simulation.

In this study, we utilise GDPR-compliant data obtained from mobile applications with explicit consent from device users to share their locations. Data privacy is a critical concern, and even anonymised data can pose risks of revealing individual identities. To mitigate privacy risks, the study strictly avoids publishing trajectories that could be traced back to specific individuals. The proposed model also holds the potential to provide representative, synthetic, and privacy-preserving mobility data.

The proposed model's methodology demonstrates strong transferability by leveraging large-scale geodata and survey data. While the model gives promising results in Sweden, applying the model in diverse settings requires fine-tuning parameters to account for variations in data collection, infrastructure, and behaviour. Adaptations to local conditions and thorough validation are essential to ensure relevance and utility.

Appendix 1

Data

The proposed model combines two data sets for synthesising activity-travel plans: the mobile application data and the Swedish national travel survey (2011–2016) (Trafikanalys). We also use Demographic statistics areas (DeSO zones) (SCB), and the building locations (lantmateriet) as auxiliary data. Below is a brief description of each data set.

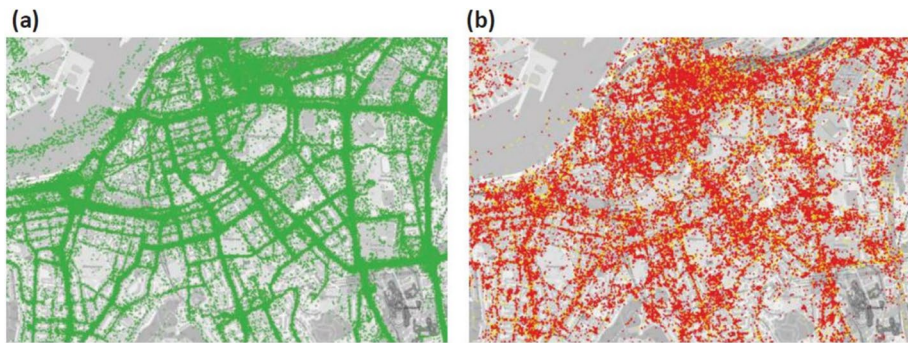


Fig. 8 Geolocations from mobile application data in Gothenburg (Liao et al. 2024). **a** Movement points. **b** Stationary points

Mobile application data

Mobile application data consists of anonymised geolocation information collected from location-enabled applications on individuals' smartphones. These data include unique anonymised device identifiers along with latitude, longitude, and timestamp information. This kind of data features extensive spatial and population coverage at a relatively low cost. However, despite providing detailed mobility traces from a substantial number of devices, the data lacks socio-demographic information about the device users to protect their privacy. To enable further analysis, relevant socio-demographic attributes, such as home locations and employment status, can be inferred from the spatio-temporal mobility characteristics.

This study uses a dataset sourced from a diverse set of mobile apps used by adult smartphone users in Sweden.¹ The data covers seven months in 2019 (from June to December) with approximately 25 million daily GPS records. In the raw data, we have around 4.6 billion records from 1 million devices. Mobile application data captures geolocations of individuals in motion (Fig. 8a) and stationary (Fig. 8b) at the moment when they interact with certain phone applications.

However, the mobile application geolocations have varying sampling frequency across users, periods, and geographic regions. This variability is primarily due to users' preference for using certain geolocation-enabled applications across different places at different times. This intermittent data collection leads to sparsity in observation and also introduces sampling bias among users, complicating the analysis of human mobility patterns.

Swedish National Travel Survey

The Swedish National Travel Survey (Trafikanalys) collects detailed information on individuals' socio-demographic characteristics, e.g., age, gender, employment status, household structure, and self-reported one-day travel diaries, including trip purposes, travel modes, distances, and timing. The survey period is between 2011 and 2016, with around 38,000 participants aged 6 to 84 years. For our study, we only use weekday activity-travel patterns, excluding data from holidays and weekends.

¹ <http://www.pickwell.co/>.

The data preprocessing involves removing records with incomplete or inaccurate trip details, e.g., trip origin and destination information. This step ensures that only participants with complete activity-travel patterns for a weekday are included in our analysis. To have a similar population with the mobile application data, we also exclude respondents below 18. After preprocessing, the data contains approximately 15,000 individuals for weekdays. Although the preprocessing step results in a reduced sample size and a further sparsity in the survey's geographic coverage across Sweden, the data offers a detailed snapshot of individuals' daily activities, providing valuable insights into their routines and travel behaviour.

Demographic statistics areas (DeSO zones)

Demographic statistical areas (DeSO zones) (SCB) obtained from Statistics Sweden are a nationwide division. DeSO zones are formed by dividing municipal boundaries into smaller areas, each with a population minimum of 700 and a maximum of 2,700 inhabitants. Across Sweden, there are a total of 5,984 DeSO zones. For each zone, the data contains various statistics, i.e., the population size, the population size by age groups, the number of employees and students, etc. We use statistics to enhance the representativeness of the results.

Building data

Building data (lantmateriet) derived from property registers provide information on more than 8.6 million buildings across Sweden. This comprehensive dataset includes building locations, footprints, and usage types (detached houses, apartments, workplaces, schools, etc.). Due to factors such as GPS inaccuracies and user movement, geolocations obtained from the Mobile Application Data (MAD) can be dispersed around actual activity locations. We employ building data to determine the precise activity locations.

Stop detection

We detect stops from individuals' geolocation trajectories, represented as (*id*, *lat*, *lon*, *time*). Stop refers to a location where an individual stays for a certain duration. These stops are critical for analysing individuals' activity-travel patterns. Various algorithms are available for stop detection, e.g., infostop (Aslak and Alessandretti 2020), DBSCAN (Ester et al. 1996), and scikit-mobility (Pappalardo et al. 2022). For this study, we follow a previous study (Liao et al. 2024) using the infostop algorithm due to its resilience to noise in data, efficiency with extensive data, and capability to conduct analyses across multiple users simultaneously (Aslak and Alessandretti 2020).

The infostop algorithm operates with parameters, $r_1, r_2, t_{min}, t_{max}$. The algorithm begins by detecting stops when an individual remains within a distance r_1 for a minimum duration t_{min} and a maximum time difference t_{max} between consecutive records. Subsequently, the algorithm clusters these stops spatially, with a parameter r_2 . Each detected stop is recorded as a tuple (*id*, *lat*, *lon*, *start*, *end*).

Table 4 shows the employed parameters in the infostop algorithm. We selected the parameters after testing several alternatives. The 15-minute dwell-time threshold strikes a balance between the accuracy and robustness of mobile phone GPS data. Shorter durations are more prone to noise, such as momentary traffic delays or artificial pauses due to sparse

or inconsistent signal records. These thresholds help filter out spurious or ambiguous stop events and ensure the identification of meaningful activity locations in large-scale mobile datasets.

Temporal-score approach for inferring primary activities

We design a scoring approach to infer primary activity locations, integrating the temporal rules with the empirical evidence into the methodology. This approach considers variations in activity patterns at different observation times rather than treating all hours equally while assessing the likelihood of locations.

We first calculate weights using travel survey data to deduce temporal rules. These weights represent the ratio of hourly activity participation for a specific activity type (e.g., home or work) to the overall activity participation in the travel survey. For home activities, weights are assigned during predefined nighttime hours from 6:00 pm to 7:59 am. For work activities, we extend weight calculations across all hours of the day to accommodate varying work schedules, including late hours, thereby offering more flexibility. Figure 9 illustrates these weights.

We then calculate a score for each activity location using the activity records with the stay duration information. Individuals can have abundant activity records in activity locations. To determine the score, each hour spent at a location is multiplied by its corresponding hourly weight. We then sum up these weighted hours and calculate the location's score. For example, if an individual has two activity records at a location from 9:00 pm to 5:59 am and from 7:00 pm to 3:59 am, each hour spent during these periods is weighted according to the specific hourly weights and summed to produce the location's overall score.

Let $W = (w_1, w_2, w_3, \dots, w_{24})$ represent the hourly weights where each w_i is associated with a specific hour of the day. For a given activity location A , let $A = (a_1, a_2, a_3, \dots, a_i)$ denote the set of activity records at this location, with each record a_i containing the hours h_i spent during each period of activity. The score S_A for location A can be calculated as follows:

$$S_A = \sum_i a_i \sum_j w_j \cdot h_j \quad (8)$$

Table 4 The infostop algorithm parameters and set values (Liao et al. 2024)

Parameter	Definition	Unit	Value
r_1	The maximum allowable distance for roaming between two points in the same stop	meter	30
r_2	The maximum distance between two stops	meter	30
t_{min}	The minimum duration of a stay	min	15
t_{max}	The maximum time difference between two consecutive records in the same stop	hour	3

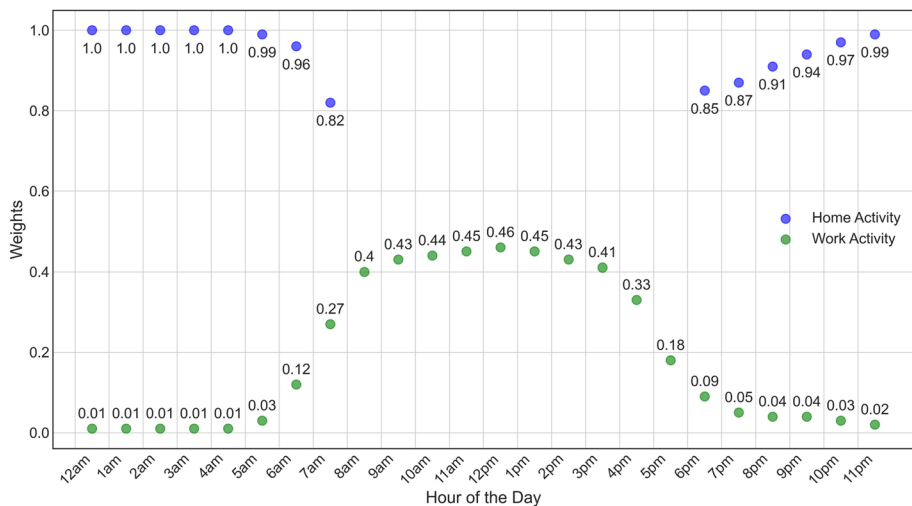


Fig. 9 Hourly weights based on activity participation in the travel survey. The sum of weights can exceed one due to rounding errors or some individuals engaging in multiple activities within an hour

Searching twin travellers

To find twin travellers for the individuals in the mobile phone dataset, we first divide individuals in the mobile phone data and the travel survey data into groups who share similar demographic and activity-travel attributes. We use the geographical context of residence, employment status, and average trip distance to reflect variations in daily travel patterns. Prior research highlights their influence on daily activity and travel patterns. For example, employed individuals typically follow regular commuting routines, whereas those not in employment tend to travel more locally (Rasouli et al. 2015). Similarly, trip-length distributions reflect the extent of individual mobility (Pappalardo et al. 2015). Additionally, the geographical context of residence, particularly urban versus rural settings, plays a significant role in shaping individuals' daily activity and travel patterns (Matz et al. 2015).

The variables used for grouping are detailed in Table 5 and include residential region, urban density of residence, employment status, average trip distance, and commuting distance. For mobile phone users, we infer these variables using the processes described earlier; for survey respondents, they are computed directly from survey responses.

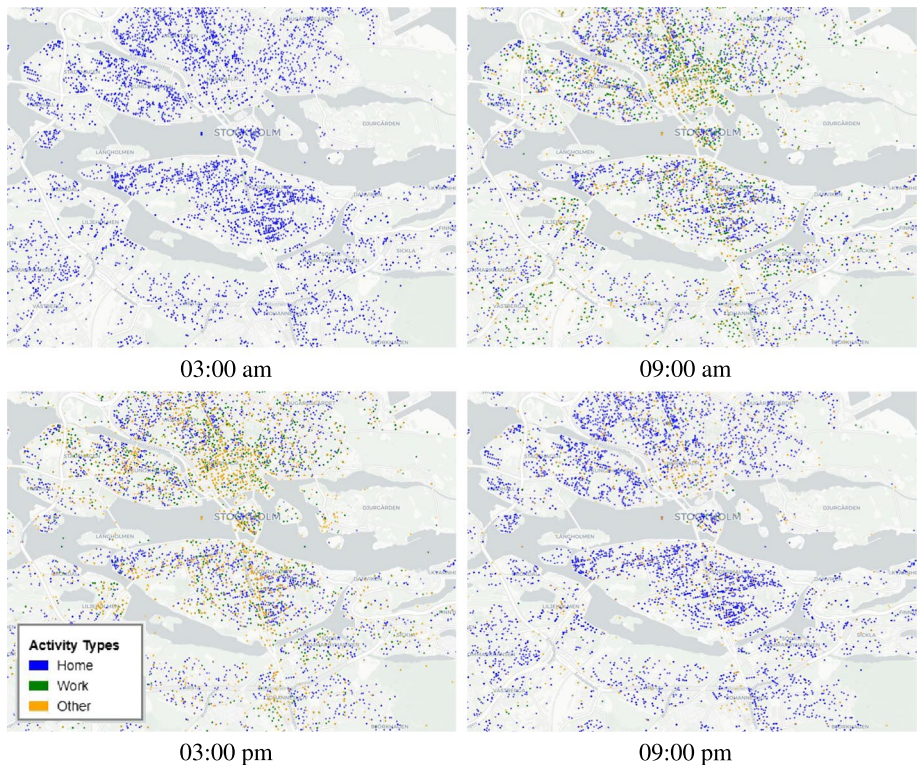
We define residential regions by the three greater regions in Sweden: Norrland, Svealand, and Götaland. Urban density, derived from DeSO zones, categorises residential locations as high or low. High urban density indicates central urban locations, while low urban density represents areas outside major population concentrations or urban areas (SCB). Employment status has two categories: employed (Yes) and unemployed (No). For the travel survey data, we based on survey responses, while for the mobile phone data, employment status is inferred based on the presence of a workplace location.

We also categorise trip distances into two levels: low and high. We calculate the average trip distances between home and all observed activity locations, as well as commuting distances between inferred home and work locations, using Euclidean trip distances. An individual's average trip distance that exceeds the median of all average trip distances is

Table 5 Variables used for categorizing travellers (section “Searching twin travellers”).

Order	Variable	Levels
1	Residential region	Svealand Götaland Norrland
2	Urban density of their residence*	Low High
3	Employment status	Yes No
4	Average trip distance	Short (< 4.3 km) Long ($\geq$ 4.3 km)
5	Commuting distance	Short (< 7.9 km) Long ($\geq$ 7.9 km)

Order column shows the variables’ sequence in the step-wise grouping algorithm for the travel survey participants.
* Defined by the DeSo zones’ statistics

**Fig. 10** Distribution of activities by time of day in central Stockholm

classified as long; otherwise, it is classified as short. The same median-based classification applies to commuting distances.

In the mobile phone dataset, all attributes are applied directly to group individuals. However, given the limited sample size of the travel survey data, we implement a step-wise approach for grouping the travel survey participants to prevent overly sparse groups. We begin grouping individuals by region and then sequentially by urban density, employment

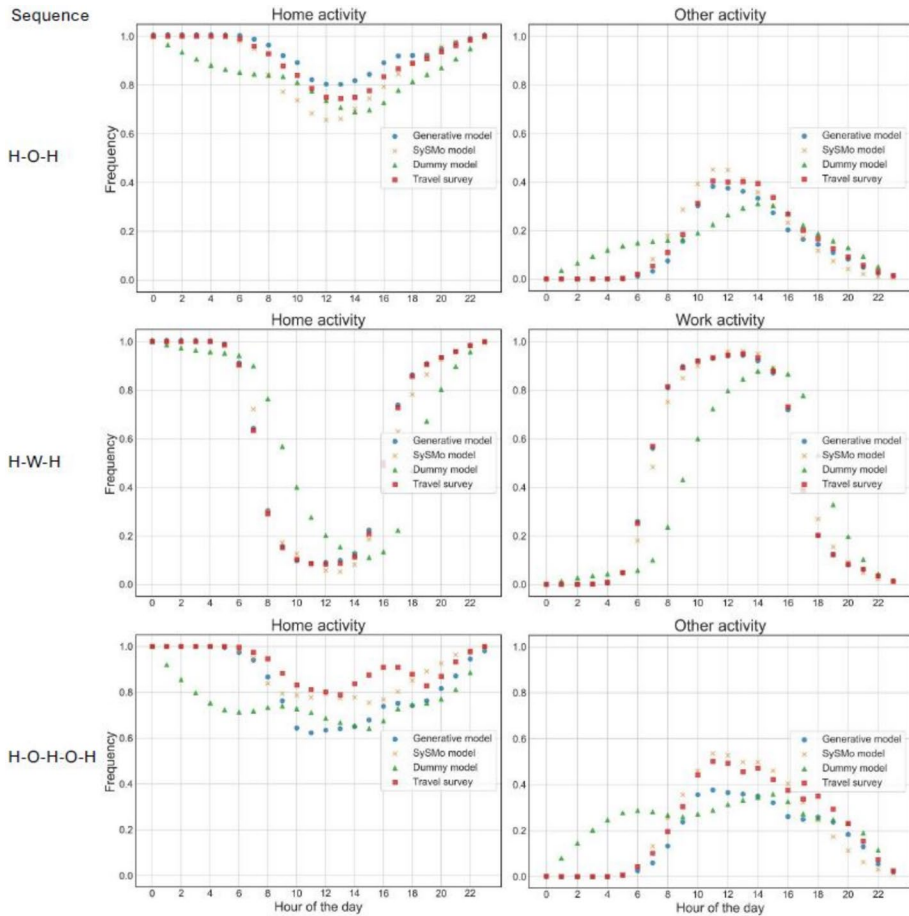


Fig. 11 Aggregated temporal activity engagement patterns by activity type and activity sequence. Rows represent the three most frequent daily activity sequences (from top to bottom): home–other–home (H–O–H), home–work–home (H–W–H), and home–other–home–other–home (H–O–H–O–H)

status, average trip distance, and commuting distance. At each step, we evaluate the size of the groups, i.e., how many participants belong to each group. If a group includes at least 50 individuals, we further subdivide the group based on the next attribute in the sequence. This sequence prioritises attributes that create larger, more stable groupings. While most participants are unaffected by the grouping order, a small number require relaxing lower-level criteria by stepping back one level in the hierarchy. This approach aims to have sufficient travellers in each group and ensures that mobile users are not repeatedly matched with survey participants in sparsely populated groups.

Distribution of activities by time of day

Figure 10 illustrates the distribution of activity participation by activity type and time of day in central Stockholm. The home activity locations are distributed across Stockholm, and the

activity participation occurs throughout the day. The work activities predominantly occur during daylight hours. The other activities are widespread on the map and show increased intensity during the daytime, with some extending into the night hours.

Temporal activity patterns by activity sequences

Figure 11 presents the joint distribution of activity sequences and temporal participation patterns. The rows show the three most frequent daily activity sequences, while the columns display aggregated temporal activity engagement by activity type. This joint representation provides a more comprehensive comparison of activity participation across models and underlying data. The generative model captures the patterns observed in SySMo model and the survey data. While the dummy model captures overall trends, it performs noticeably worse in reproducing detailed behavioural patterns.

Acknowledgements This research is funded by the Swedish Research Council Formas (Project Number 2018-01768). The authors acknowledge Sonia Yeh for her intellectual contributions to the study's methodology. Additionally, the authors sincerely thank Jorge Gil for providing the mobile phone application data and Mattias Rydström and Diana Salim for their master thesis as valuable prior work.

Author contributions Conceptualization: Çağlar Tozluoğlu (Ç.T.), Yuan Liao (Y.L.), Frances Sprei (F.S.); methodology: Ç.T., Y.L., F.S.; software: Ç.T., Y.L.; validation: Ç.T.; data curation: Ç.T., Y.L.; writing - original draft: Ç.T.; writing - review & editing: Ç.T., Y.L., F.S.; project administration: F.S.

Funding Open access funding provided by Chalmers University of Technology. This research is funded by the Swedish Research Council Formas (Project Number 2018-01768).

Data availability We followed the Chalmers Institutional Review Board (IRB) guidelines in line with the Swedish Ethical Review Act (2003:460) for research involving humans and the General Data Protection Regulation (GDPR) 2016/679. Given the nature of the data analyzed, the study was exempt from ethical review under the Swedish Ethical Review Act.

Declarations

Ethics approval The authors declare that this work does not involve the use of social media data or experimentation with animals. We followed the Chalmers Institutional Review Board (IRB) guidelines in line with the Swedish Ethical Review Act (2003:460) for research involving humans and the General Data Protection Regulation (GDPR) 2016/679. Given the nature of the data analyzed, the study was exempt from ethical review under the Swedish Ethical Review Act.

Generative AI and AI-assisted technologies in the writing process During the preparation of this work, the authors used ChatGPT in order to improve language and readability, with caution. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflict of interest The authors declare no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the

copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahas, R., Silm, S., Järv, O., Saluveer, E., Tiru, M.: Using mobile positioning data to model locations meaningful to users of mobile phones. *J. Urban Technol.* **17**(1), 3–27 (2010)
- Alessandretti, L., Aslak, U., Lehmann, S.: The scales of human mobility. *Nature* **587**(7834), 402–407 (2020)
- Alexander, L., Jiang, S., Murga, M., González, M.C.: Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. C: Emerg. Technol.* **58**, 240–250 (2015)
- Arentze, T.A., Timmermans, H.J.P.: ALBATROSS: a learning based transportation oriented simulation system. EIRASS (2000)
- Aslak, U., Alessandretti, L.: Infostop: scalable stop-location detection in multi-user mobility data. arXiv preprint [arXiv:2003.14370](https://arxiv.org/abs/2003.14370) (2020)
- Bassolas, A., Ramasco, J.J., Herranz, R., Cantú-Ros, O.G.: Mobile phone records to feed activity-based travel demand models: MATSim for studying a cordon toll policy in Barcelona. *Transp. Res. A Policy Pract.* **121**, 56–74 (2019)
- Bowman, J.L., Ben-Akiva, M.E.: Activity-based disaggregate travel demand model system with activity schedules. *Transp. Res. Part A Policy Pract.* **35**(1), 1–28 (2001). [https://doi.org/10.1016/S0965-8564\(99\)00043-9](https://doi.org/10.1016/S0965-8564(99)00043-9)
- Bricka, S., Bhat, C.R.: Comparative analysis of global positioning system-based and travel survey-based data. *Transp. Res. Rec.* **1972**(1), 9–20 (2006)
- Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. *Nature* **439**(7075), 462–465 (2006)
- Chen, C., Bian, L., Ma, J.: From traces to trajectories: how well can we guess activity locations from mobile phone traces? *Transp. Res. C: Emerg. Technol.* **46**, 326–337 (2014)
- Chin, K., Huang, H., Horn, C., Kasanicky, I., Weibel, R.: Inferring fine-grained transport modes from mobile phone cellular signaling data. *Comput. Environ. Urban Syst.* **77**, 101348 (2019)
- Deming, W.E., Stephan, F.F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Stat.* **11**(4), 427–444 (1940)
- Endres, D.M., Schindelin, J.E.: A new metric for probability distributions. *IEEE Trans. Inf. Theory* **49**(7), 1858–1860 (2003)
- Ester, M., Kriegl, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *kdd*, vol. 96, pp. 226–231 (1996)
- Ficek, M., Kencl, L.: Inter-call mobility model: a spatio-temporal refinement of call data records using a gaussian mixture model. In: 2012 Proceedings IEEE INFOCOM, pp. 469–477. IEEE (2012)
- Fu, X., Zhang, Y., Ortúzar, J.D.D., Lü, G.: Activity-travel pattern inference based on multi-source big data. *Transp. Rev.* **45**(1), 26–48 (2025)
- Golledge, R.G.: Spatial behavior: a geographic perspective. Guilford Press, New York (1997)
- Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L.: Understanding individual human mobility patterns. *Nature* **453**(7196), 779–782 (2008)
- Grujić, N., Brdar, S., Osinga, S., Hofstede, G.J., Athanasiadis, I.N., Pljakić, M., Obrenović, N., Govedarica, M., Crnojević, V.: Combining telecom data with heterogeneous data sources for traffic and emission assessments—an agent-based approach. *ISPRS Int. J. Geo Inf.* **11**(7), 366 (2022)
- Hilgert, T., Heilig, M., Kagerbauer, M., Vortisch, P.: Modeling week activity schedules for travel demand models. *Transp. Res. Rec.* **2666**(1), 69–77 (2017)
- Hörl, S., Balac, M.: Synthetic population and travel demand for Paris and île-de-France based on open and publicly available data. *Transp. Res. C: Emerg. Technol.* **130**, 103291 (2021)
- Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C., Pujolle, G.: Estimating human trajectories and hotspots through mobile phone data. *Comput. Netw.* **64**, 296–307 (2014)
- Huang, H., Cheng, Y., Weibel, R.: Transport mode detection based on mobile phone network data: a systematic review. *Transp. Res. C: Emerg. Technol.* **101**, 297–312 (2019)
- Janzen, M., Vanhoof, M., Smoreda, Z., Axhausen, K.W.: Closer to the total? long-distance travel of French mobile phone users. *Travel Behav. Soc.* **11**, 31–42 (2018)
- Jiang, S., Ferreira, J., Gonzalez, M.C.: Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore. *IEEE Trans. Big Data* **3**(2), 208–219 (2017)
- Kitamura, R.: A model of daily time allocation to discretionary out-of-home activities and trips. *Transp. Res. Part B: Methodol.* **18**(3), 255–266 (1984)

- Kung, K.S., Greco, K., Sobolevsky, S., Ratti, C.: Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS ONE* **9**(6), e96180 (2014)
- lantmateriet. GSD Property Map by Lantmateriet. <https://www.lantmateriet.se/sv/geodata/vara-produkter/pruduktlista/fastighetsindelning-nedladdning-vektor/>. Accessed 20 Feb 2020
- Lee, R.J., Sener, I.N., Mullins, J.A.: An evaluation of emerging data collection technologies for travel demand modeling: from research to practice. *Transp. Lett.* **8**(4), 181–193 (2016)
- Liao, Y., Tozluoğlu, Ç., Sprei, F., Yeh, S., Dhamal, S.: Impacts of charging behavior on BEV charging infrastructure needs and energy use. *Transp. Res. Part D: Transp. Environ.* **116**, 103645 (2023)
- Liao, Y., Gil, J., Yeh, S., Pereira, R.H., Alessandretti, L.: The effect of limited mobility on the experienced segregation of foreign-born minorities. *npj Sustainable Mobility and Transport*, 29 (2025). <https://doi.org/10.1038/s44333-025-00046-4>
- Matz, C.J., Stieb, D.M., Brion, O.: Urban-rural differences in daily time-activity patterns, occupational activity and housing characteristics. *Environ. Health* **14**, 1–11 (2015)
- Miller, H.J.: Activities in space and time. In: *Handbook of Transport Geography and Spatial Systems*, pp. 647–660. Emerald Group Publishing Limited (2004)
- Moeckel, R., Huang, W.-C., Ji, J., Llorca, C., Moreno, A.T., Staves, C., Zhang, Q., Erhardt, G.D.: The activity-based model ABIT: modeling 24 hours, 7 days a week. *Transp. Res. Proc.* **78**, 499–506 (2024)
- Namazi-Rad, M.-R., Tanton, R., Steel, D., Mokhtarian, P., Das, S.: An unconstrained statistical matching algorithm for combining individual and household level geo-specific census and survey data. *Comput. Environ. Urban Syst.* **63**, 3–14 (2017)
- Ogle, J., Guensler, R., Elango, V.: Georgia's commute Atlanta value pricing program: recruitment methods and travel diary response rates. *Transp. Res. Rec.* **1931**(1), 28–37 (2005)
- Österreicher, F., Vajda, I.: A new class of metric divergences on probability spaces and its applicability in statistics. *Ann. Inst. Stat. Math.* **55**(3), 639–653 (2003)
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.-L.: Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**(1), 8166 (2015)
- Pappalardo, L., Simini, F., Barlacchi, G., Pellungrini, R.: Scikit-mobility: a python library for the analysis, generation, and risk assessment of mobility data. *J. Stat. Softw.* **103**(1), 1–38 (2022)
- Rasouli, S., Timmermans, H., Waerden, P.: Employment status transitions and shifts in daily activity-travel behavior with special focus on shopping duration. *Transportation* **42**, 919–931 (2015)
- Rodr, A.B., Burrieza-Gal, J., Jos, J., Castro, I.P., Wilby, M.R., Cantu-Ros, O.G., et al.: Using app usage data from mobile devices to improve activity-based travel demand models. *IEEE Trans. Big Data* **10**, 633–643 (2024)
- Sadeghinassr, B., Akhavan, A., Wang, Q.: Estimating commuting patterns from high resolution phone GPS data. In: *ASCE International Conference on Computing in Civil Engineering 2019*, pp. 9–16. American Society of Civil Engineers Reston (2019)
- Sadilek, A., Kautz, H., Silenzio, V.: Modeling spread of disease from social interactions. In: *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 6, pp. 322–329 (2012)
- SCB. Demographic Statistical Areas (DeSO). <https://www.scb.se/en/services/open-data-api/open-geodata/de-so--demographic-statistical-areas/>. Accessed 20 Nov 2019
- Schönfelder, S., Axhausen, K.W.: Activity spaces: measures of social exclusion? *Transp. Policy* **10**(4), 273–286 (2003)
- Song, C., Koren, T., Wang, P., Barabási, A.-L.: Modelling the scaling properties of human mobility. *Nat. Phys.* **6**(10), 818–823 (2010)
- Stalla-Bourdillon, S., Knight, A.: Anonymous Data V Personal Data-false Debate: an Eu Perspective on Anonymization, Pseudonymization and Personal Data. *Wis. Int'l LJ* **34**, 284 (2016)
- Thierry, B., Stanley, K., Kestens, Y., Winters, M., Fuller, D.: Comparing location data from smartphone and dedicated global positioning system devices: implications for epidemiologic research. *Am. J. Epidemiol.* **193**(1), 180–192 (2024)
- Tozluoğlu, Ç., Dhamal, S., Yeh, S., Sprei, F., Liao, Y., Marathe, M., Barrett, C.L., Dubhashi, D.: A synthetic population of Sweden: datasets of agents, households, and activity-travel patterns (2023a). <https://data.mendeley.com/datasets/9n29p7rmn5/2>. Accessed 20 May 2024
- Tozluoğlu, Ç., Dhamal, S., Yeh, S., Sprei, F., Liao, Y., Marathe, M., Barrett, C.L., Dubhashi, D.: A synthetic population of Sweden: datasets of agents, households, and activity-travel patterns. *Data Brief* **48**, 109209 (2023)
- Tozluoğlu, Ç., Liao, Y., Sprei, F.: Potential of e-bikes to replace passenger car trips and reduce greenhouse gas emissions. *J. Cycl. Micromobility Res.* **2**, 100043 (2024)
- Trafikanalys. The swedish national travel survey (2016). Accessed 20 Feb 2021 <https://www.trafa.se/en/travel-survey/travel-survey/>

- Van de Kerckhove, W., Mohadjer, L., Krenzke, T.: A weight trimming approach to achieve a comparable increase to bias across countries in the programme for the international assessment of adult competencies. In: JSM Proceedings, Survey Research Methods Section, pp. 655–666. American Statistical Association, Alexandria (2014)
- Vanhoof, M., Lee, C., Smoreda, Z.: Performance and sensitivities of home detection on mobile phone data. Big data meets survey science: a collection of innovative methods, pp. 245–271 (2020)
- Widhalm, P., Yang, Y., Ulm, M., Athavale, S., González, M.C.: Discovering urban activity patterns in cell phone data. *Transportation* **42**, 597–623 (2015)
- Wischer, T., Cik, M., Fellendorf, M.: Graph supported mode detection within mobile phone data trajectories. *Transp. Res. Rec.* **2677**(3), 18–32 (2023)
- Wojtusiak, J., Nia, R.M.: Location prediction using GPS trackers: can machine learning help locate the missing people with dementia? *Internet Things* **13**, 100035 (2021)
- Woodard, D., Nogin, G., Koch, P., Racz, D., Goldszmidt, M., Horvitz, E.: Predicting travel time reliability using mobile phone GPS data. *Transp. Res. C: Emerg. Technol.* **75**, 30–44 (2017)
- Yabe, T., Jones, N.K., Rao, P.S.C., Gonzalez, M.C., Ukkusuri, S.V.: Mobile phone location data for disasters: a review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **94**, 101777 (2022)
- Yu, Q., Li, W., Yang, D., Zhang, H.: Mobile phone data in urban commuting: a network community detection-based framework to unveil the spatial structure of commuting demand. *J. Adv. Transp.* **1–15**, 2020a (2020)
- Yu, Q., Zhang, H., Li, W., Song, X., Yang, D., Shibasaki, R.: Mobile phone GPS data in urban customized bus: dynamic line design and emission reduction potentials analysis. *J. Clean. Prod.* **272**, 122471 (2020)
- Yu, Q., Zhang, H., Li, W., Sui, Y., Song, X., Yang, D., Shibasaki, R., Jiang, W.: mobile phone data in urban bicycle-sharing: market-oriented sub-area division and spatial analysis on emission reduction potentials. *J. Clean. Prod.* **254**, 119974 (2020)
- Zhang, H., Song, X., Long, Y., Xia, T., Fang, K., Zheng, J., Huang, D., Shibasaki, R., Liang, Y.: Mobile phone GPS data in urban bicycle-sharing: layout optimization and emissions reduction analysis. *Appl. Energy* **242**, 138–147 (2019)
- Zhang, H., Chen, J., Li, W., Song, X., Shibasaki, R.: Mobile phone GPS data in urban ride-sharing: an assessment method for emission reduction potential. *Appl. Energy* **269**, 115038 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Çağlar Tozluoğlu¹  · Yuan Liao^{1,2}  · Frances Sprei¹ 

✉ Çağlar Tozluoğlu
caglar.tozluoglu@chalmers.se

¹ Department of Space, Earth and Environment, Chalmers University of Technology, Gothenburg, Sweden

² Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark